AWS CERTIFIED SOLUTION ARCHITECT -2020 [SAA-CO2] BY NEAL DAVIS

GETTING STARTED- SECTION 2

**Identify And Access Management (IAM) overview:**

* IAM User:
* An IAM user is an entity that represent a person or service.
* Can be assigned: an access key ID and secret access key for programmatic access to the AWS API, CLI, SDK, and other development tools.
* A password for access to the management console.
* By default users cannot access anything in your account.
* Root account has full administrative permission and these cannot be restricted.
* IAM users can be created to represent applications and these are known as “service accounts”.
* ARN (Amazon Resource Name) uniquely identifies the user across the AWS.
* Password policy can be defined for enforcing password length, complexity etc.
* IAM Groups:
* A collection of users and have policies attached to them.
* Collection of users and have policies attached with them
* A group is not an identity and cannot be identified as a principle in an IAM policy
* Use groups to assign permissions to users.
* Use the principal of least privilege when assigning permissions.
* Cannot have group over group.
* IAM Roles:
* Roles are assumed by trusted entities and can be used for delegation.
* Roles are created and then “assumed” by trusted entities and define a set of permission for making AWS service request.
* With IAM Roles you can delegate permission to resources for users and services without using permanent credentials e.g. (username and password).
* IAM users or AWS services can assume a role to obtain temporary security credentials that can be used to make API calls.
* No credentials associated with a role (password or access keys).
* IAM policy:
* List of permission for users that is allowed to do in aws service.
* Policies are documents that define permission and can be applied to the users, groups and roles.
* Policy documents are written in JSON (key value pair that consists of an attribute and a value).
* All permissions are implicitly denied by default.
* The most restrictive policy is applied
* The IAM policy simulator is a tool to help you understand test and validate the effects of access control policies.
* The condition element can be used to apply further conditional logic.
* IAM Access Key:
* A combination of an access key and a secret access key.
* Users can be given access to change their own keys through IAM policy (not from the console)
* You can disable a user’s access key which prevents it from being used for API calls.
* These can be used to make programmatic calls to AWS when using the API in program code or at a command prompt when using the AWS CLI or the AWS PowerShell tools.
* You can create, modify, view and rotate access key.
* When created IAM returns the access key ID and secret access key
* The secret access key is returned only at creation time and if lost a new key must be created.
* Ensure access key and secret access key are stored securely

IAM Console Password:

* A password that the user can enter to sign into interactive sessions such as the AWS Management Console.
* You can allow IAM users to change their passwords by disabling the option for all users and using an IAM policy to grant permissions for the selected users.

Multi-Factor Authentication:

* A secure password
* Virtual MFA /Physical MFA

AWS Security Token Services (STS):

* The aws security token service is a web service that enables you to request temporary limited privilege credentials for IAM users or for users that you authentic (federal users).
* All Aws STS requests go to a single endpoints at <https://sts/amazon.com>.
* All regions are enabled for STS by default but can be disabled.
* The region in which temporary credentials are required must be enabled.
* Credential always work globally .
* **Regional edge caches** are Cloud Front locations that are deployed globally, close to your viewers. They're located between your origin server and the POPs—global **edge** locations that serve content directly to viewers. ... If the files are in the **cache**, Cloud Front returns them to the user.
* Cloud Front **Edge locations** are connected to the **AWS** Regions through the **AWS** network backbone - fully redundant, multiple 100GbE parallel fiber that circles the globe and links with tens of thousands of networks for improved origin fetches and dynamic content acceleration.
* (Virtual Private Cloud)**VPC** is a logically isolated section of the AWS cloud where you can launch your own resources .Within VPC you can create your own networks using own IP ranges. VPC sits in a region. We can launch our resources in Sub net.
* **Classless Inter domain Routing:** Classless **inter domain routing** (CIDR) is a way to categorize Internet **Protocol** (IP) addresses for allocation to hosts and more efficient **routing**. CIDR represents the IP address and its subnet mask with a single number.
* Internet Gateway attached to your VPC and that allows us to **access the outside world**. For example, we can request out to and Internet Web site through the internet gateway.04312

EC2 - Chapter 02

* EC2 is the elastic compute cloud and it's an elastic service that allows you to launch compute resources on the AWS cloud. Each Instance has an operating system some storage and a certain amount of virtual hardware associated with it including CPU, ram and network capabilities.
* Limits determine what we are allowed to run in the AWS cloud. The limit can be extendable if we can contact to aws. By default for instance we can only run a certain amount of a resource on an AWS.
* If we go on the IAM dashboard and launch an instance, it's called launching an **On-Demand Instance** and we don't get any discounts but it's very flexible you can just launch these and terminate them whenever you want to.
* In **reserved Instances,** we can purchase a reserved instance and you can choose the *amount of time* that you want to reserve this for. The options are 12 months or 36 months so one year or three years and this means that you're committed to paying for these instances for that period of time.

Now in return for this commitment, you get a big discount can be up to 75 percentage of the on-demand price. This can be used for steady-state workloads that you know you want to run for that period.

**Spot Request Instance,** where spot requests what you are doing is purchasing some spare capacity from AWS at a much-discounted rate. Aws often have some capacity that's not being used and they'll give you the opportunity to use some of that capacity at a discounted rate.

* So people use this for instance for large compute jobs that are very expensive normally and they really need that sort of reduction in cost.
* **IMPORTANT::**This is useful for large computing jobs that are doing some kinds of batch processing. When Aws need that capacity Aws can shut down or take this power back. So whole using these instances, make sure we have saved/stored somewhere too.

**Note:** On Demand Instances is expensive than Spot Request Instance.

**Saving Instances** is similar to reserved Instances but it gives you the option to choose different compute models such as EC2 and Fargate which is a container service.

* Another option that we have is dedicated hosts a **dedicated host** is a physical server in EC2 that you can use the capacity of and it's dedicated for your use.

Dedicated hosts provide dedicated hardware and they give you full visibility of sockets and cores and targeted instance placement.

* That means even though EC2 is typically multi-tenant so you're sharing the underlying hardware with other customers. If you use a dedicated host you're not sharing the underlying hardware.
* **Scheduled Reserved Instances,** this is where we can purchase some capacity on a recurring schedule. If we run certain workloads on a recurring basis but not constantly then scheduled reserved instances is the best option.
* **Capacity Reservation** is actually for reserving some capacity to ensure it's available for you when you actually need to run some resources. It can be useful if we have a highly critical workload and we need to make sure that we can run it at a certain time.
* Our PC we can be connected into EC2 instances using **Internet Gateway.** One Internet gateway attached to One VPC that enable to access the internet. From the internet gateway, we can connect to **EC2 instances**.
* If your instance has more than one network interface it won't pick up a new dynamic IP address while disassociating the elastic address.
* If your instance has a secondary private IP address with an elastic IP associated with it, it also won't pick up a new dynamic public address.
* **Security groups** are firewalls that are applied at the instance level. So they actually look for traffic that's going into and out of the plastic interface is connected to our instances. We can have multiple instances in the same security group and we can have multiple security groups applied to instances.

The key factor about Security group is stateful that means if you allow an inbound connection the response traffic is always allowed.

* (ssh ec2-user@publicIpAddress -i keypairfilepastehere) //On Windows

**NOTE:** Sometimes we have to modify the key pair downloaded file’s security permission for example we have to modify file’s permission.

* The allocation of inbound traffic on **ICMP protocol** from the security groups means that any instance that's a member of this security group will be able to send ICMP data to other instances that are members of the security group but no one else will.
* **Instance metadata** is data about your instance that you can use to configure or manage the running instances.

<http://169.254.169.254/latest/meta-data/>

* **Instance User Data :** Instances user data is available at : http://169.254.169.254/latest/user-data/

#!/bin/bash

yum update -y

yum install httpd -y

systemctl start httpd

systemctl enable httpd

cd /var/www/html

echo "A demo page running on EC2 Instances in the AWS cloud environment" >index.html

* **Elastic IP** address is a static public address, which is charged if not used and associated with a private IP address on the instance.
* **Private Subnet:** Every instance will have a private address whether it's in a public subnet or private subnet. It’s always going to have one and the address is going to be from the range assigned to our VPC.
* **Retained** when the instance is stopped, used in public and private subnets.
* The key difference between Elastic and public IP address is that the elastic IP address is static (same IP) whereas the public IP address is dynamic (changeable IP).The other key difference is you can move an elastic IP address between instances and adapters.

Elastic network adapters are just additional interfaces that you can attach to your instance.

* **Internet Gateway** **job** is to change the private and elastic IP address into public address and send it out to the internet and vice-versa.
* A **Bastian Host** is literally an instance in a public subnet that you use to then jump to your private subnet. Some people called it as jump post.
* Each subnet can be associated with one and only one availability zone.

\*\*\*\*\*\*Connect from windows with Agent forwarding ---MAKE REVISION\*\*\*\*\*\*

* Nat stands for network address translation which really is the process of taking a private IP address and translating it to a public address so that you can speak on the internet.

The network load balancer is the one that operates at the connection level.

#!/bin/bash

yum update -y

yum install httpd -y

systemctl start httpd

systemctl enable httpd

cd /var/www/html

aws s3 cp s3://dctlabsa/names.csv ./

aws s3 cp s3://dctlabsa/index.txt ./

EC2NAME =’cat ./names.csv|sort -R|head -n 1|xargs’

sed “s/INSTANCE/$EC2NAME/” index.txt > index.html

From the **Exam point of View**:

Very high performance and very low latency and lots of connection, we have to think about the network load balancer.

* **Application Load Balancer:**
* Operates at the request level
* Routes based on the content of the request (layer 7)

#!/bin/bash

yum update -y

yum install httpd -y

systemctl start httpd

systemctl enable httpd

cd /var/www/html

aws s3 cp s3://dctlabsa/names.csv ./

aws s3 cp s3://dctlabsa/index.txt ./

EC2NAME =’cat ./names.csv|sort -R|head -n 1|xargs’

sed “s/INSTANCE/$EC2NAME/” index.txt > index.html

cp index.html orders.html

* For application load balancer we have to choose **lambda function**.
* Auto Scaling is a service that enables you to automatically launch instances you want to launch and auto scaling takes care of launching those instances and also making sure that the number of instances is always maintained.
* Failing an instances mean that they are failing a health check.
* We cannot modify the **launch configuration** that means if this launch configuration is attached to an auto-scaling group and we need to **change** something like the **instance type** you want to use you can't make this modification it's important for the exam.
* **Fleet Composition**: We could adhere to launch template or we can combine purchase options and instances. On this way, we can **mix** On Demand instances with spot instances and multiple instance types.
* **Auto scaling** architecture includes high availability and fault tolerance.
* The behavior is slightly different for each type of load in the **sticky session**.
* We cannot have two listener with port 443.
* An Instance can be in multiple target groups.

Virtual Private Cloud (VPC):

DNS hostname enabled or disabled options:

**Explanation:**

This attribute indicates where instances with public IP addresses get a corresponding public DNS hostname. So we remember when we've launched EC2 instances into public subnets. The DNS hostname is provided which contains the IP address as a part of the name. So as long as you set this to true those type of hostnames will be generated.

* We cannot have as much security groups within an instance that means we can only have five security group within an instance.
* Network ACL works as a firewall in a subnet level. The main difference is network ACL is associated with subnets not with instances.
* VPC pairing is a way that you can set up a network connection so that you can send traffic between V.P. seats using private IP addresses so we can use the site blocks of our VPC is to send traffic directly to other pieces and those can be in the same account or they can be in a different account and they can even be in a different region.
* If an application demands low latency and you see some exam questions and one option is direct connect and another is VPN then in the case where you need to ensure low latency you definitely go for direct connect it is much more costly than a VPN and it takes some time as well it can take months to provision.
* For the exam just remember that if you need to connect using private that's so private IP address spaces globally and it's a use case for Direct Connect e.g. it's something along the lines of low latency high bandwidth dedicated network connection but you need to better connect over private IP address spaces globally then you might be looking to a direct connect gateway as the answer.
* Just remember when you create a new subnet it will be associated with a default route table (VIP Exam).
* So you might see an exam question that says that a malicious IP address has been identified. You need to very quickly block that specific IP or that specific range of IPs and you could do that from a network ACL but not a security group so you can't do a security groups and network ACL is the first line of defense because traffic going into the subnet hits the network ACL first and then the security group gets hit because that's associated with the instance which is already within the subnet.

AWS Direct Connect (Exam)

* Whenever you see questions that mention creating a hybrid cloud network connection it's probably going to be direct connect because it's a private link. It requires a large network link and you'd use it when you have lots of resources and services being provided on AWS and you need that high bandwidth low-latency connection.
* So always remember if there's a very quick timeline for deployment like you need to get something set up within two weeks or a month. Direct connects, probably not going to be provisioned in that time frame it could be two or three months .We then have to use direct connect plus VPN.
* So if you see an exam question that says that you need to enable encryption on a direct connect connection .It might be a VPN so you need a virtual private gateway and a customer Gateway. VPN cloud.

Section 6: Amazon Route 53:

**Note ::**

You should look out for questions that talk about locking down distribution of content localizing web pages or serving Web pages in different languages that kind of thing.(Geo Location Routing Policy)

Section 7: Amazon S3 and IAM Roles Overview

* Amazon S3 is a simple storage service and it’s an object based Stored System.
* S3 is a global services.
* The maximum file size for Amazon S3 objects is 5 terabytes.
* In Amazon S3 you get eventual consistency for overwrite PUTS and DELETES.
* Transfer Acceleration speeds up data uploads by using the Cloud Front network.
* You can mimic the hierarchy of a file system by creating folder in your buckets.
* S3 is a global service but buckets are created within a region. Data is never replicated outside of that region unless you configure it (e.g. through CRR).
* Amazon S3 Glacier Deep Archive is the most cost-effective option for these requirements as the data retrieval time is 24 hours.
* Cloud Front is a content delivery network and is ideal for this use case as it caches the content around the world, provides a single endpoint address, and uses a single source for the videos.
* AWS Lambda is not a valid origin for Amazon Cloud Front.
* **Role** is an identity in which we can assign permissions through policy and then you’re able to assume that role.

AWS S3 and Cloud Front

So whenever exam questions come up just remember S3 is the object storage system and you use HTTP verbs (methods) to access a public endpoint, EBS is for block-based storage, hard disk drives, solid state drives, operating systems, and you must be in the same availability zone as your instance. And then EFS is when you're looking at file systems which you might want to mount from multiple EC2 instances within different Availability Zones and even on premises clients over a VPN so use cases for EFS include home directories, shared corporate directories, database backups, big data analytics Just remember you cannot issue block level commands or format file systems with NFS mounted file system that's already being shared from somewhere else.

Choosing Access Control Options

We then have the ACLs and the ACLs apply at the bucket and object level and have less granularity especially in terms of the permissions that can be assigned and the uses that they can be assigned to so the account level not the user level.

Multipart Upload

The main thing to remember for the exam is that it's recommended to use multipart upload when your files are 100 megabytes or more in size. Typically in a production scenario you would use one of the SDK and do it programmatically but as you can see even if you use the client will automatically break files up for you.

Query String Authentication

The pre signed URL can be valid for a maximum of seven days because the signing key you use in the signature calculation is valid for up to seven days.

Transfer Acceleration

The end point that you would use if you wanted to accelerate your uploads is called transfer acceleration.

* So you end up using a different endpoint to access and your data is uploaded to a cloud front edge location then forwarded onto your bucket.
* So just remember if you have latency and your uploads are latency sensitive or that time sensitive so you need to get that data into your bucket quickly and you will benefit financially or by some kind of benefit to the business if you get that data out quickly that it might be worth spending the extra money to ensure that your upload is quicker.
* Couple of other notes about transfer acceleration as I said before it kind of cannot be disabled only suspended. Also you are actually only charged if there is a benefit but as far as I know that means if you get some benefit whether it's a small benefit or not you'll be charged. So you've got to work out if you do get a small benefit is it cost effective and just also remember for the exam that you must use the correct endpoint to get the advantage. So you have to use the extra accelerate endpoints.

Static Websites

You can also put a cloud front distribution in front of S3 so that it can cache your content around the world.

Multiple Versioning

So as AWS web page mentions versioning is a means of keeping multiple variants of an object in the same bucket. So what that means is that every time you upload and change a objects the previous version and the current version are retained.

* The first thing to know is that buckets can be in one of Free states. So beautiful a bucket does not have versioning enabled. You can then enable it and then you can suspend it. You cannot return back to the original state.
* What will happen is if we delete the newer version and head here and open the file again we've turned back to the previous version and then the other thing we can do is we can delete this file and then if we click on Show we can see that the file the original file is here and then there's a delete marker and I can delete the delete marker and now the file is back again so that's how versioning works.

Cross Region Replication:

 CRR enables you to asynchronously copy objects between buckets that are in different AWS regions so AWS already replicates your data between geographically distant Availability Zones but you might want to copy it to even greater distances across regions. Versioning should be enabled to initiate cross region replication.

* You also might do it so that you can minimize the latency in other words the distance between your users and your data.
* It might also be for operational reasons so you might have computer clusters in different regions they could be processing some data and you want that information to be locally available for those compute.

Lifecycle Management:

How many zones are there in you know S3 standard IA or one zone IA. But you do need to know this because it might ask you a question about what the redundancy is or you might have a constraint in a question that says that you need to ensure that your data is replicated between at least three Availability Zones which storage tier should you use.

And so it might sway you one way or another. Also understand the minimum storage duration which applies to certain tiers and then also you can see here you pay per gigabyte retrieved with some tiers but not with others. So just get an understanding of what the charges are I'd get an understanding of which constraints apply to each tier and how they're designed and architected.

* It tells you what these supported transitions are and what the unsupported transitions are. So you should be aware of these just understand you know which what you can move from where to where.

S3 Encryption:

We then have the service side encryption with client provided keys SSE-C. So in this case the encryption still takes place up here in the cloud.

Requester Pay:

It doesn't support anonymous requests. It doesn't also support bit coin or SOAP requests. Now also the request must include certain parameters in the header of the request so that specifies Information that is used then to ensure that that user is charged for those requests and that data transfer back in the bucket.

Server Accessing Logging:

Another security feature of S3 is server access logging with server access logging you can log the requests that are made to S3.So we can do a couple of ways we can enable it for the console. You can enable it programmatically and it gives you a bit of information on the page here about the format of the logs now. The interesting thing is that its best effort server log delivery so you're pretty much says here that you will get a log or you maybe you won't get a log. It’s not even guaranteed.

Object Lock:

So this is another tool in your arsenal so you've got versioning, you've got multifactor authentication delete protection and then you've got object lock delete protection and then you've got object lock.

S3 Select and Glacier Select:

We have S3 select what you can do is use a sequel expression to look inside the zip file and then only extract the data that you actually need so that can obviously reduce the amounts of data that you need to download and it can also reduce the cost because your if you're pulling this out of a W.S. then that outbound data transfer fee is going to be reduced as well.

Cloud Front Overview:

So what's the difference between an edge location and a regional edge cache?

Well basically what this means is that the edge location has a cache and a regional edge location has a cache. Now there are a lot more edge locations than there are regional edge caches. But the regional edge caches have a bigger cache. So what that means is if your data is being frequently accessed then it's going to get cached at edge locations and is going to be constantly pulling that down so the time to live is going to keep getting refreshed as the object is being used.

So an origin is simply where your data comes from. So that could be an S3 bucket that could be an 3 bucket configured as a static web site or that could be a EC2 instance or an EC2 instance sat behind an application load balancer. So it is recommended to use EC2 instances behind ALB because obviously you want to ensure availability of the actual origin.

Cloud Front and Static Websites:

So what we're going to do instead is we're going to set up a bucket policy that only allows the cloud front distribution to connect to the static web site and there's some authentication using what's called an origin access identity which is a type of user which you can create on Cloud Front.

So that secures our content to make sure that it only comes free Cloud Front now sometimes creating Cloud Front distributions can take quite some time.

Cloud front with ALB and EC2 Custom Origin

And what this means is you can create AWS Lambda functions that execute closer to your users and say you know the example that says here is you can create a trigger that causes the function's execute when Cloud Front receives a request from a viewer for a specific cache behavior you set up for your distribution. So this is another way of getting your logic closer to your users as well as your content.

This is definitely something that can come up in the exam so it's worth just reading up on Lambda Edge just to understand exactly what it is.

Amazon S3 Storage Class: Exam Cram

So basically for everything except S3 standard and there's also a **retrieval fee** that's associated with these four tiers here. Make sure you know those because they can come up in exam questions. There are quite a lot of exam questions on S3 so it is quite a big subject and you do need to really understand it well.

Amazon S3 Transfer Acceleration: Exam Cram

There's no difference in the security is just as secure as uploading directly to S3 and you're only charged if is a benefit in transfer times compared to just uploading straight to S3.You have to enable this on the bucket and you can then disable it you can only suspend it.

Amazon S3 Performance: Exam Cram

You use byte range fetches and you use retry of requests for licensee sensitive applications and you can combine S3 and EC2 in the same AWS region so that's what you should do. Also it's recommended to use transfer acceleration to minimize latency caused by distance. Now again this is one where there's quite a lot of information to read up on here to understand the difference.

Best Practices for performance.

**Please check the training notes and the AWS documentation.**

I'll put a link in. There's definitely exam questions coming up asking about bite range fetches or where the answer is to use byte range feature. So you need to understand what that is.It's really just a way of creating multiple parallel requests for your data to improve performance.

Amazon Cloud Front: Exam Cram

There's also a special type of user called The Origin access identity (OAI) and that can be used to restrict access to content in an S3 bucket.

If you use an OAI you can restrict users so they cannot access the content directly using theS3 url. They have to connect to cloud front.

Definitely an exam question watch out for this one. So if you have an S3 bucket you can use OAI to make sure that no one can circumvent cloud front and go directly to the S3 bucket but you can't do it with an EC2 Origin.

ELASTIC BLOCK AND FILE SYSTEM

EBS vs. EFS

 The key differences between EBS and EFS are:

* One EC2 instance accessing one volume.
* We can have multiple volumes attached to the same EC2 instance but you can't have multiple EC2 instances attached to a single volume so you can't share volumes between computers.
* You can detach it from one computer and then attach it to another computer but they can't both access it at the same time with EFS you can have multiple computers accessing the same file system so they can be reading and writing to the exact same file system and you can have them across multiple Availability Zones whereas with EBS the volume and the EC2 instance must be in the same availability zone.
* You can even have an on premises client connecting into your VPC over a VPN and accessing the EFS file system.
* Now one note point is that Linux only for EFS. You can't connect a Windows instance to EFS. You can pretty much run whatever you want on your EC2 instance with a block storage system.

EBS vs. Instance State

* So there are some use cases where people do want to use these you can use them for some kind of temporary data , that's perhaps distributed across instances and replicated in some other way and where you need that high level of performance.
* So with an ephemeral volume (Instance store) you can't stop your instance so you'll lose your data you can restart your instance though.
* And you can't detach an ephemeral volume from one instance and attach it to another like you can with an EBS volume.
* So those are just a few facts about ephemeral storage. If you see ephemeral that's an instance store and that means that the data is not persistent.

EBS VOLUME TYPES

So you do need to know this stuff for the exam.

* Remember, snapshots are stored on Amazon S3 which stores data within a region, not an AZ.
* The general purpose SSD and the provisioned IOPS so we can see that the use case for the general purpose
* SSD is most workloads so they said this is a good balance of price to performance, recommended for most workloads it's good for system boot volumes good for virtual desktops and low latency interactive apps and dev. test environments.
* On the other hand your provisioned IOPS is about mission critical low latency so this is really high performance .So it's more about business critical applications large database workloads that kind of thing then you have throughput optimized hard disk drives you are looking here at lower cost and the lowest cost is the cold but that's very low performance so you really only use it for some very specific scenarios where all you care about is cost something is not very important to you with the HDD one rather than the SDD one.

So this is more about throughput so low cost but pretty good throughput. So there might be some use cases that you want to use this for.

* And again you might want to just have a look at what's the cost differences you can come down and see that it gives you 500 megabytes a second there vs. a thousand for general purpose and provisioned IOPS. So there's quite a bit more performance in the provisioned IOPS SSD drive in terms of throughput but then again the cost of this is going to be so much lower and you could even potentially bundle a bunch of these disks together in a raid volume to aggregate the throughput.

So just familiarize yourself with this table and understand the different use cases and also the limitations so you can't use even these as a boot volume.

Now you remember back here we did have an option magnetic standard which could be a big volume so you can still use that but it's not even on this table anymore. So really it's just being deprecated away.

Another thing to note maximum volume size is the same across all of these 16 tibibytes whereas the max IOPS is significantly different so much higher on these SSD than it is on these HDD. Another thing that you might know is the minimum size is so a gigabyte or gigabyte for here 500 here.

Launch instances with Multiple EBS Volumes

We could choose provisioned IOPS and really it would just be a performance difference. So I don't need that for this lab but you could just choose that if you wanted to so notice a difference had the delete on termination is enabled for the volume but it's not enabled for any additional volumes this is a key exam question sometimes you'll find something come up where it mentions that an instance has been terminated it had a volume and it had a couple of additional volumes added to it you know what data is going to be accessible and you know that the additional volumes by default will be available because they're not goanna be deleted.

Create volume fro snapshots

Well, we'll take a snapshot of the EBS volume and then we'll create a volume from that snapshot. And what we're actually leveraging there is the fact that the snapshots are stored on S3 which is a regional service. So even though EBS volumes only exist in a specific availability zone, because our snapshots are on S3 we're then able to restore the snapshot to any availability zone within the region.

Working With Snapshots:

* So you'll want to look up deleting an Amazon EBS snapshot and the key thing is here is that it says if you make periodic snapshots of a volume the snapshots are incremental so that's what I just explained .And that means that only the blocks on the device have changed after your last snapshot are saved in the new snapshot. Now in most backup systems that would mean that you need all of those snapshots in order to recover.
* Another tool that we have is data lifecycle manager and this is a way that you can automate the lifecycle of snapshots. It’s a relatively new service but it is several months old so it will be featuring on the exam now.
* Now there's another thing about snapshots that you need to understand to take a fully consistent snapshot. Sometimes referred to as a crash **consistent snapshot**. You must stop right to the volume. So that means that to ensure that you have a fully consistent snapshot you need to stop any activity that's going on the volume. Of course that's very hard for a root volume when the operating system is running on it. So ideally you would shut down your instance and you would then take a snapshot of it.

EBS Performance:

This happens in AWS at the operating system level so in other words you configure this through your operating system you would attach multiple volumes and then you'd configure one of these types of RAID like RAID zero or RAID one for your operating system so it is actually important to know this for the exam. I would at least know the difference between RAID zero which is striping and RAID one which is mirroring your discs, and you need to understand what the performance and redundancy benefits are of each of these different types of RAID.

And that's all here. So you can look that up here. Another one is we have nested RAID. So this is where you put two types of RAID together such as RAID 10.So that's another one that's potentially going to come up on the exam. Understand RAID 10, understand RAID zero, and RAID one. And just remember that you need EBS optimized instances or at least instances with a decent amount of network bandwidth at least 10 gigabits per second.

Amazon EFS Access Control:

EFS is a file-based storage system accessed over NFS. You can attach thousands of instances from multiple AZs to the same file system.

You can control access to files and directories with POSIX-compliant user and group-level permissions.

* Now there are two performance modes with AFSC you've got general-purpose, which is suitable for most use cases,
* And then you've got Maximo where you can scale to higher levels of aggregate throughput and operations per second. Attribute Fornier file system, something that might push you to use Maximo instead of general purpose.

So just watch out where exam questions might ask for a particular performance.

Amazon FSx:

FSx for Windows File Server provides NTFS file systems that can be accessed from up to thousands of compute instances using the SMB protocol.

Now for the SAA-C02 exam and we can have few questions on my exam.

* So this is the FSx for Windows file server. And what you have is you can connect your Windows based application. So for instance you might have two instances running Windows but you could also have on premises clients running Windows connected over a VPN or direct connect connection and you can then talk to a file system and that file system supports the native Windows file system features such as NTFS file systems and ACLs, shadow copies and user quotas amongst other features and has access over the server message block or SMB protocol.
* The key thing that we'll notice here first off is that we have an S3 bucket. This is key to know for the exam so you'll need to identify when to use FSx for Windows versus NSX for Lustre.

**Key exam point here.**

* The thing to remember is whenever you see S3 being mentioned and it looks like an FSx use case it's gonna be NSX for Lustre.
* And if it's SMB or if it's gonna be a Windows based application then it's going to push you to FSx for Windows.

AWS Storage Gateway -File Gateway:

When you get exam questions which type of storage Gateway you should be using or whether you should be using a storage gateway at all.

* So the first one is a file gateway. So with a file Gateway the storage Gateway here is actually an appliance.
* Gateway offers SMB your NFC based access to data in S3 with local caching and it supports stays free different tiers of all classes of S3 storage so S3 standard II or one zone II.
* It also supports Linux clients connecting to the Gateway using interface versions free and full and Windows clients connecting to the Gateway using SMB versions 2 and free maximum file size of an individual file is 5 terabytes that's the free limit.
* You need to know all the facts for the exam. So just make sure you understand the different types of Gateway and when they're going to be used.

Amazon Elastic Block Store – Exam Cram

Watch out for exam questions here.

* I've noticed instance store coming up a bit more often now and the type of scenario where it's really good is you might need some really high I/O some really high storage performance and the data is replicated across multiple instances.
* I'm not going to go through all of this but just make sure you're aware of the differences.

Amazon EFS and FSX –EXAM Cram

If you're doing that SAA-C03 version of the exam you definitely need to know Amazon FSx, there could be a few questions on the exam and I recommend therefore going and just doing a bit more reading up on our training notes.

AWS-Storage Gateway – Exam Cram

* All data is encrypted using SSL and all data stored by the time gateway in S3 is encrypted server side with Amazon S3 managed encryption keys.
* So again remember those two last points those apply to all of the gateways. So you always have SSL encryption for data in transit and then for data at rest you're going to house as free server side encryption.

Section-9: CONTAINERS:

 Dockers is a software provider who really started the recent movement towards containerization and ECS is compatible with Dockers the application runs quickly and reliably from one computing environment to another a docker container image is a lightweight standalone executable package of software that includes everything needed to run an application code runtime system tools system libraries and settings.

* Using the Fargate launch type, you can run your containerized applications without the need to provision and manage the backend infrastructure. After you register your task definition, Fargate launches the container for you. **Fargate** users charges a higher per-hour fee than ECS and EKS users
* With the EC2 launch type, you can run your containerized applications on a cluster of Amazon EC2 instances that you manage.

Elastic Container Registry:

In this lab we're going to cover the elastic container registry.

* ECR is a fully managed docker container registry and it's where you store the images of your containers so you store images in the registry and then those images are pulled down based on your task definition and they're pulled down to create your task or your actual container itself.

Amazon ECS Auto Scaling:

It uses a new resource type quarter capacity provider and that's associated with an auto scaling group.

* So what this means is you're able to scale not just the tasks the number of containers you're running automatically but you're now able to automatically scale the number of cluster instances the hosts on which your clusters or your easiest tasks actually run. It leverages two new features ones called manage scaling and once called managed instance termination protection.

Section -10 SERVERLESS:

Key point in Exam:

* Supporting Programming languages in AWS Lambda are: (Net Core, Go, Java, Node.js, Python, and Ruby).
* Concurrency is the concurrent executions of functions which is running so over multiple executions of function running in running at any one time. The concurrency limit on AWS account is 1000.
* Go various differences of EC2, ECS (EC2 Launch Type), ECS (Fargate Launch Type) and Lambda. Generally the Question can be: What is the most scalable solution? What is the most operationally efficient solution?

AWS Lambda:

The poll based services. So when I say poll based, what I mean is that AWS Lambda is polling Dynamo DB. So this is the complete opposite of the last lab where you had S3 and S3 was sending an event notification when something happened in a bucket. So it was S3 that was responsible for triggering Lambda to execute its code.

 It was S3 that was responsible for triggering Lambda to execute its code. In this scenario Dynamo DB is a stream based service as is Kinesis and SQS. And so Lambda polls Dynamo DB .It would poll Kinesis, it would poll SQS, so the same thing for those three services and what this means is the configuration for the event source mapping is on Lambda not on the service itself.

Key Point to know exam question you need to know that difference.

API Gateway Overview:

* I think it was late 2018 so probably starting to appear on the exam now. They used to just offer rest API. But now API Gateway also offers web socket API as well as REST API. Our HTTP based adhere to the rest protocol and use standard methods it's worth dicing that with API Gateway.
* Another thing that's worth noting is that lambda and API Gateway form the app facing parts of the AWS service infrastructure so this sometimes comes up in exam questions where they might ask you to identify or they might ask you what would go along with one of these services as part of the app facing part of the AWS service infrastructure so that's it.
* These are basically your front-end services for your application.
* Let's just head back to the diagram now and we've covered a bit of this on the web page that we just looked at. But again you've got your front-end so you've got your clients could be a mobile client a service could be a Web site, could be IOT, could be pretty much anything on the Internet that's going to talk REST.
* So there are three different types of endpoints. Originally API gateway came out with the edge optimized endpoint only. And therefore if your user base is global then you can reduce your latency for requests that come from anywhere around the world. So that was a very good solution if you are a global company or you have a global user base.
* And if you remember the differences the regional API is it deployed in the region. The Edge optimized API uses cloud front. That’s the globally distributed content delivery network, and then the private API is for when you only want to access your API from within your VPC from your own private IP ranges. So we can leave that as regional and click import.
* We can specify the time to live and the authorization it's just to select that you then have throttling this sometimes comes up on the exam. Your account level throttling right is 10000 requests per second with a burst of 5000. Now you can see that this stage actually has 10000 requests per second.